DOCUMENT RESUME

ED 314 428 TM 014 146

AUTHOR	Echternacht, Gary
TITLE	Interpreting Test Scores for Compensatory Education Students. ERIC Digest.
INSTITUTION	American Institutes for Research, Washington, DC.; ERIC Clearinghouse on Tests, Measurement, and Evaluation, Washington, DC.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
REPORT NO	EDO-TM-89-5
PUB DATE	Dec 89
CONTRACT	RI88062003
NOTE	3p.; For document from which this Digest was extracted, see TM 014 144.
PUB TYPE	Information Analyses - ERIC Information Analysis Products (071)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Achievement Tests; *Administrator Role; Administrators; *Admission Criteria; *Compensatory Education; Educational Assessment; Elementary Secondary Education; *High Risk Students; Objective Tests; *Program Evaluation; Scores; Selection; Standardized Tests; *Test Interpretation; Test Use
IDENTIFIERS	ERIC Digests; Fact Sheets; Selection Tools

ABSTRACT

Following the rules and regulations of compensatory education programs requires the use of objective measures in selecting students for programs, assessing their progress, and monitoring the program's quality. Administrators must remember the following four points when they use test scores for compensatory education students: (1) test scores alone should not be used to select students; (2) good programs select students through several assessment tools rather than just one; (3) administrators should not use out-of-level tests; and (4) the use of tests for lower grade levels may be less frustrating for some students, but will not give an accurate picture of the content learned or the real meaning of the scores. It is also important to understand that the term "grade level" in testing does not relate to how well students perform in the classroom. When students score at grade level, about half of their peers score higher and about half score lower. Administrators must also differentiate the degree of error in individual and group scores. Interpretations are most sure when administrators consider district averages, followed by building averages, classroom averages, and individual students' scores. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED314428

ERIC Clearinghouse on Tests, Measurement, and Evaluation Digest EDO-TM-89-5 December 1989

Interpreting test scores for compensatory education students

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Gary Echtemacht, Educational Testing Service

To follow the rules and regulations of compensatory education programs correctly, you must use objective measures when you select students for programs, assess their progress, and monitor the program's quality. Because you have this pressure to use standardized test scores, you should make sure that you use the tests correctly.

In this digest, I point to four practices that administrators often mistakenly follow when they use test scores:

- using test scores alone to select students for programs,
- giving out-of-level tests,
- misinterpreting grade-level, and
- failing to differentiate the degree of error in individual and group scores.

Although these practices may not be widespread, they are serious.

Don't use test scores alone to select students for programs

Program regulations for Chapter 1 require that you select students by using objective measures. In addition, state departments of education sometimes impose other requirements -- for example, a program can serve only students who score below the 40th percentile rank or all students who score below the 20th percentile rank.

These requirements often lead administrators to select students on the basis of test scores alone because

- the requirements are stated in terms of test scores, and
- when program monitors review programs, they appraise them in terms of state and federal regulations.

Nevertheless, you should not make a decision about an individual student by using a test score by itself. It is acceptable to use test scores to make decisions in a sequence of assessments, but it is unacceptable to use test scores by themselves in a sequence of one assessment. You are unfair to students if you simply say that all students who score below the 40th percentile rank are in the program and all who score above the 40th percentile rank are ineligible.

You must remember that test scores are neither completely reliable nor valid indicators of academic performance. For example, if students take an equivalent form of a test at different times, their scores will change somewhat. This unreliability is important for those whose scores are near the cut-off score for selection because if you administer the same test a second time, some students who previously scored **below** a cut-off may score **above** the cut-off a second time.

Similarly, reading tests give you only general measures of reading ability. Some students may be good readers in certain content areas, yet they may score poorly on a given test because the reading passages in that test do not include the content areas they know.

Good programs select students by using several assessment tools, rather than just one. Although the regulations do not explicitly state other requirements, they do allow you to use additional assessment tools in selecting students. Ask your state director how you can best use other assessment tools, such as report card grades, results of other tests, and systematic teacher assessments obtained through questionnaires.

Some common methods for using multiple assessments are:

- selecting students who score below prescribed cut-offs on both your district's standardized test and another state-mandated test;
- using your district's standardized test to identify a pool of possible participants, then using either a teacher-completed questionnaire or report card grades to select students from the pool;
- using a systematic method for obtaining teachers' judgments about students' needs in order to identify a pool of possible participants, then using a standardized test to select students from the pool; or
- using the standardized test to identify a pool of students, then creating a study team to select students from the pool and carefully documenting the study team's process.

Don't give out-of-level tests

Out-of-level testing occurs when you give a standardized test to students who are at a different grade level than the one for which the test is designed. In some cases, school officials use out-of-level tests in compensatory programs because those students are behind their peers and in-level testing is frustrating for them. Administrators who follow this practice believe that somehow it is more valid to give those students tests designed for lower grade levels.

TM 014146

American Institutes for Research, 3333 K Street, NW Suite 300, Washington, DC 20007 (202) 342-5060

While out-of-level tests may be less frustrating to some students, the scores obtained from them are also less valid because

- the content for out-of-level tests does not represent the content taught in the classroom,
- the scale that test publishers use to link different test levels is loaded with error,
- there are no norms for out-of-level tests,
- scores obtained on tests of different difficulty are not comparable, and
- when obtained, out-of-level scores appear to be too low.

Although in-level test scores are more reliable in the middle than at the high- and low-score ranges, they are quite reliable in placing students at the high or low end of the scale. For example, with a reasonable degree of assurance, we can say that a student who scores at the 10th percentile rank is most likely a low-achieving student. What we are less sure about is whether the student is at the 10th percentile rank or the 15th percentile rank. Either way, we are reasonable in concluding that the student is low achieving.

You should use tests at the grade levels for which they are specified by the test publisher. Generally, the content of grade-level tests will represent what is taught in regular classrooms at the specified level.

If your compensatory program is good, it will be closely coordinated with instruction in the regular classroom. Since the purpose of compensatory education is to help students succeed in the regular classroom, using in-level tests will help you in the coordination.

Understand the term "grade-level"

Generally, when school personnel say that certain students perform at grade-level, they mean that those students can learn material at about the same rate and quality as others in the same class. The implication is that students who don't perform at grade-level have significantly more difficulty in class than their peers. Accordingly, when students are labeled as working below grade-level, the implication is that they may not have the aptitude, maturity, or interest to do the work that others in the same class are doing. This interpretation of students' abilities is made by relatively few people.

In contrast, in the testing arena at grade-level has a different meaning. When students score at grade-level, their scores are at the 50th percentile rank. It means that about half of their peers score higher and about half score lower. In testing, at grade-level does not relate to how well students perform in the classroom. Therefore, when you review students' scores, you must consider that, by definition, many students score below grade-level.

Historically, the term grade-level has been important in the politics of compensatory education. Proponents of compensatory education programs have always said that those programs were underfunded because many students who performed below grade-level did not receive program services. In this case, performing below grade-level was defined as scoring below the 50th percentile rank. While it is true that compensatory education may be underfunded and, I believe, is an important part of schooling, it is inappropriate to use the term grade-level in the true testing-related sense.

Since most people use the term grade-level in the general sense, you can either avoid using grade-equivalent test scores or develop a range of scores that indicate satisfactory achievement in the classroom. You may also think of average performance on a test as being between the 23rd and the 77th percentile rank.

Differentiate the degree of error in individual and group scores

Administrators tend to interpret differences in test scores in one of two ways. First, they may think that a difference of one or two percentile rank points is an important difference. Secondly, they may think that a difference of ten points shows that the test is unreliable. Few administrators can differentiate the degree of error in individual and group scores.

An individual test score is just that -- the score that an individual student receives on a test. A group score is the average of several individual scores. For example, the average score of third graders at Horace Mann Elementary School is a group score.

In general, individual scores have more error in them than group scores do. The error in an individual score is largely a function of the test's standard error that is described in the publisher's technical manual. For most of the tests given in elementary and secondary schools, the standard error is about 2.5 raw score points. This means that about 95% of the time, we would expect the scores for individual students to fall within a range of 10 raw score points. That is not particularly reassuring, but it is exactly why we need to use multiple measures for selecting students and why for most of the tests we use we should be a little skeptical of individual test scores and cautious in interpreting differences.

The error in group scores largely depends on the size of the group. Once you have a group of about 30 scores, the magnitude of the errors decreases. By the time you average all the scores for your school district, you can regard the results as accurate as long as there is not some systematic bias operating for most everyone in the district.

You can be confident of your interpretation when you consider score averages of large groups. For instance, if when you consider a group of 55 scores, the score average changes one or two percentile rank points, then that is an important change. If you consider averages based on fewer cases, you must be more cautious. You can be more or less confident of average scores depending on the level. There is a definite hierarchy in the strength of your interpretations. Your interpretations are most sure when you consider district averages, followed in order by building averages, classroom averages, and finally individual students' scores.

This publication was prepared with funding from the Office of Educational Research and Improvement (OERI), U.S. Department of Education, under contract R-88-062003. The opinions expressed in this report do not necessarily reflect the position or policy of OERI or the Department of Education. **Permission is granted to copy and distribute this ERIC/TM digest.**